A specialized metabolic network selectively modulates *Arabidopsis* root microbiota

Ancheng C. Huang,[1#] Ting Jiang,[2,3,4#] Yong-Xin Liu,[2,3] Yue-Chen Bai,[5,6] James Reed,[1] Baoyuan Qu,[2,3] Alain Goossens,[5,6] Hans-Wilhelm Nützmann,[1,†] Yang Bai,[2,3,4*] Anne Osbourn[1*]

Affiliations:

[1]Department of Metabolic Biology, John Innes Centre, Norwich Research Park, Colney Lane, Norwich NR4 7UH, UK.

[2]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Science, Beijing, China.

[3]CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences (CAS), Beijing 100101, China.

[4]University of Chinese Academy of Sciences, College of Advanced Agricultural Sciences, Beijing 100039, China.

[5]Ghent University, Department of Plant Biotechnology and Bioinformatics, Ghent 9052, Belgium.

[6]VIB Center for Plant Systems Biology, Ghent 9052, Belgium.

[†]Current address: Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK.

[#]These authors contributed equally to this work.

*Correspondence to: anne.osbourn@jic.ac.uk or ybai@genetics.ac.cn

Abstract

Plant specialized metabolites have ecological functions, yet the presence of numerous uncharacterized biosynthetic genes in plant genomes suggests that many molecules remain unknown. Here, we discover a triterpene biosynthetic network in the roots of the small mustard plant *Arabidopsis thaliana.* Collectively, we have elucidated and reconstituted three divergent pathways for the biosynthesis of root triterpenes, namely thalianin (seven steps), thalianyl medium-chain fatty acid esters (three steps) and arabidin (five steps). *A. thaliana* mutants disrupted in the biosynthesis of these compounds have altered root microbiota. *In vitro* bioassays with purified compounds reveal selective growth modulation activities of pathway metabolites towards root microbiota members and their biochemical transformation and utilization by bacteria, supporting a role for this biosynthetic network in shaping an *Arabidopsis*-specific root microbial community.

Short title: A triterpene cocktail for microbes.

One Sentence Summary:

The small mustard plant *Arabidopsis* produces a range of metabolites that serve to bend the microbial community around its roots to its own purposes.

Plants have evolved the ability to produce a vast array of specialized metabolites. This metabolic diversification is likely to have been driven by the need to adapt to different environmental niches (*1*). One of the key environmental factors that influences plant health and fitness is the root microbiota (*2-4*), yet mechanisms underpinning the establishment of specific root microbiota and how plants direct the microbial communities around their roots remain elusive. Plants are estimated to utilize ~20% of their photosynthesized carbon to make root-derived organic molecules, stimulating the formation of distinctive root microbiota from surrounding soil (*5-8*). However, whether and which specialized metabolites produced by plants can modulate root microbiota is not known. The presence of a substantial number of uncharacterized root-expressed biosynthetic genes in plant genomes also reveals our incomplete understanding of these communication processes.

Triterpenes are specialized plant metabolites that have important functions in plant defense and signaling. They are one of the largest and most structurally diverse families of plant natural products and many have been shown to possess antifungal and antibacterial activities, suggesting potential roles in mediating interactions between the producing plants and microbes (*9-11*). Triterpenes are synthesized via the mevalonate pathway, the first committed step being carried out by enzymes known as triterpene synthases (TTSs), which collectively are able to make a diverse array of different triterpene scaffolds (*12*). There are 13 predicted TTS genes in the genome of *A. thaliana* accession Col-0, five of which (*MRN*, *THAS*, *ABDS*, *PEN3*, *BARS*) are located in a total of four plant biosynthetic gene clusters on chromosomes 4 and 5 (Fig. 1A).

Plant biosynthetic gene clusters represent evolutionary genomic hot spots under strong selection pressure (*13*), with the potential to encode metabolites of ecological significance (*14*). Moreover, most genes within these four *Arabidopsis* biosynthetic gene clusters are root-expressed (Fig. 1B)

and the arabidiol gene cluster has been implicated in defence against the root-rot pathogen *Pythium irregulare*, suggesting that the derived metabolites might play a role in root-microbe interactions. *(15-18)*. The TTSs of the *A. thaliana* biosynthetic gene clusters appear to share a common ancestor but have functionally diverged and form a monophyletic clade that is distinct from other *A. thaliana* TTSs (fig. S1) (*15*). These clustered TTSs channel primary metabolism into specialized metabolism by converting the ubiquitous precursor 2,3-oxidosqualene into different triterpene scaffolds, specifically marneral/marnerol **M1/M2** (*19*), thalianol **T1** (*20*), arabidiol **A1** (*21*), tirucalladienol **Tr1** (*22*), and baruol **B1** (*23*) via the corresponding bi/tri/tetra-cyclic carbocations, respectively (fig. S1). Four of the 13 cytochrome P450 (CYP) genes in these cluster regions [*At5g48000* (*THAH*), *At4g15330* (*CYP705A1*), *At5g62590* (*MRO*), and *At5g36110* (*CYP716A1*) belonging to the CYP708, CYP705, CYP71, and CYP716 families, respectively] have previously been shown to convert **T1**, **A1**, **M1/M2**, and **Tr1** primarily to 3β,7β-thaliandiol **T2**, 14-apo-arabidiol **A2**, 23-hydroxy-marner-al/ol **M3/M4**, and unknown products, respectively (Fig. 2) (*18, 24*). However, numerous other uncharacterized genes with predicted functions in specialized metabolism are present in these biosynthetic gene clusters and are co-expressed with the characterized TTS and CYP genes in *A. thaliana* roots (Fig. 1B and fig. S2) (*15, 18*). Here we elucidate three biosynthetic pathways derived from the thalianol and arabidiol gene clusters and show that these major root specialized metabolites play important roles in selectively modulating *Arabidopsis* root bacteria both *in vivo* and *in vitro*, and contribute significantly to the establishment of an *Arabidopsis*-specific root microbial community.

Elucidation of the thalianol gene cluster-derived pathways

The thalianol biosynthetic gene cluster in *A. thaliana* contains four co-expressed genes, of which the functions of two (the TTS *THAS* and the CYP *THAH*) have been characterized (*15, 24*). A

second CYP (*THAO*, *At5g47990*) that belongs to the CYP705 family mediates a third step in the pathway but the precise structure of the resulting product has not been determined (*15*). A fourth gene (*THAA1*, *At5g47980*) which is co-expressed with the other pathway genes is predicted to encode an acyltransferase belonging to the BAHD family but its function is unknown (*15, 25*). We used *Agrobacterium*-mediated transient expression in *Nicotiana benthamiana* leaves to investigate the functions of the cluster genes *THAO* and *THAA1* (*26-28*). We first co-expressed *THAS* with the three other cluster genes encoding the hydroxylase (THAH), CYP705A5 (THAO) and acyltransferase (THAA1) in the leaves of *N. benthamiana* in a combinatorial fashion (table S1). We found that THAO alone could modify thalianol to give two new products, 3β,16-thaliandiol **T3** and 16-keto-thalianol **T4**, respectively when co-expressed with *THAS* (Fig. 2 and fig. S3). Co-expression of *THAO* and *THAH* with *THAS* gave 16-keto-3β,7β-thaliandiol **T5** and 16-keto-3β,7β,15-thaliantriol **T6** as the main products, instead of the desaturated thaliandiol proposed previously (Fig. 2 and fig. S3) (*15*). Furthermore, we found that *THAA1* encodes an acyltransferase that was functional only when co-expressed with all three other cluster genes, *THAS*, *THAH* and *THAO,* yielding a new product identified as 3β,7β-dihydroxy-16-keto-thalian-15-yl acetate **T7**, along with its C_{17}=C_{18} *cis* isomer ***cis*-T7** as the minor product (Fig. 2 and fig. S3). This suggests that THAA1 functions after THAS, THAH, and THAO to add an acetate group on the C15 position of the thalianol scaffold after a hydroxy moiety is installed at this position. We also identified another BAHD acyltransferase gene (*THAA2*, *At5g47950*) in close proximity to *THAA1* on chromosome five that was co-expressed with the thalianol cluster genes (Fig. 1). In contrast to THAA1, THAA2 is promiscuous and can act on the C3 hydroxy moiety of different thalianol derived compounds (**T1**-**T7**) to introduce an acetyl group (**T11**-**T17**) when co-expressed with subsets of the thalianol cluster genes (Fig. 2 and fig. S4, table S1). Our results indicate that the

four thalianol cluster genes and the nearby co-expressed gene *THAA2* are functional and yield 7β-hydroxy-16-keto-thalian-3β-15-yl diacetate **T17** when co-expressed in *N. benthamiana*.

Identification of thalianol-derived root metabolites in *Arabidopsis*

We next performed targeted metabolomic analysis to identify thalianol-derived metabolites in *A. thaliana*. Metabolic profiling identified seven major thalianol-derived products (**T1**, **T2**, **T9**, **T10**, **T18a-c**) in root extracts from the wild type Col-0 accession and the *THAS* overexpression line (*thas-oe*) that are absent in those from the *THAS* mutants (*thas-ko1* and *thas-ko2*) (Fig. 3, A to C, table S2). Mutation of *THAH* (*thah-ko*) led to accumulation of 3β,15-thaliandiol **T3** and 16-keto-thalianol **T4**, whereas mutation of *THAO* (*thao-ko*) resulted in elevated levels of 3β,7β-thaliandiol **T2** and absence of **T3** and **T4** (Fig. 3, A and B; fig. S5), indicating that THAO indeed functions as a C16 oxidase of thalianol in *A. thaliana*. We were unable to detect **T5-T7** or **T17** in either Col-0 wild type or *thas-oe* root extracts, but we did detect two compounds **T9** and **T10**, potential isomers of **T7** and **T17** (fig. S5). Compound **T10** was absent in root extracts from mutants of all thalianol cluster genes and *THAA2* (*thaa2-ko* and *thaa2-crispr*), while **T9** was still detected in those of *thaa2-ko* and *thaa2-crispr*, suggesting that **T9** and **T10** are downstream products of the thalianol gene cluster and that THAA2 may act after **T9** to form **T10** (Fig. 3, A to C and fig. S5). Targeted metabolomics analysis also revealed the presence of thalianol-derived medium-chain saturated triterpene fatty acid esters (TFAEs) in the root extracts of Col-0 and *thas-oe*, including thalianyl palmitate (16:0, **T18a**), myristate (14:0, **T18b**) and laurate (12:0, **T18c**) (Fig. 3, A and B). Compounds **T18a-b** were also detected in the roots of other mutant lines *thah-ko*, *thao-ko*, *thaa1-crispr*, *thaa2-ko* and *thaa2-crispr* except *thas-ko1* and *thas-ko2*, suggesting the presence of a branched thalianol-derived biosynthetic pathway. The identities of these TFAEs were confirmed by chemical synthesis.

Identification of missing genes for the biosynthesis of thalianin, TFAEs and arabidin

To identify the missing genes required for the biosynthesis of metabolites **T9** and **T10**, we carried out a genome-wide search for co-expressed candidate biosynthetic genes in the ATTED-II plant co-expression database (http://atted.jp/) using the four thalianol cluster genes as baits (*29*), as no biosynthetic genes in or near the gene cluster region appeared to be co-expressed. We selected eight candidates from the top 20 co-expressed genes for functional analysis in *N. benthamiana* (table S2). We identified two genes *At3g29250* (*THAR1*) and *At1g66800* (*THAR2*) that encode a pair of promiscuous oxido-reductases capable of epimerizing the C3 hydroxy moiety of **T1**-**T7** (Fig. 1, A and B, Fig. 2 and fig. S6). In this epimerization sequence, THAR1 converted the C3β hydroxy of **T1**-**T7** into the C3 ketones [**3-keto-**(**T1**-**T6**) and **T8**] while THAR2 reduced the C3-ketones into 3α alcohols [**3-epi-**(**T1**-**T6**) and **T9**] sequentially (Fig. 2, fig. S6, and table S9). Co-expression of *THAR1* and *THAR2* with the thalianol cluster genes and *THAA2* completed the biosynthesis of **T10** in *N. benthamiana* (fig. S7). The structure of **T10** was established as 7β-hydroxy-16-keto-thalian-3α-15-yl diacetate by NMR and **T10** named thalianin hereafter (table S10).

To identify genes responsible for the biosynthesis of TFAEs **T18a-c**, we screened seven *O*-acyltransferase genes (table S4) from *A. thaliana* identified based on annotation (*30*), and found that only one of these [*THAA3*, *At3g51970*; previously reported to function in sterol fatty acid ester biosynthesis (*31*)] could catalyze the formation of thalianyl palmitate **T18a** when co-expressed with *THAS* in *N. benthamiana* (fig. S8, A and B). We also generated a *THAA3* overexpression line (*thaa3-oe*) and found that THAA3 could catalyze the formation of **T18a** *in planta* since elevated levels of **T18a**, but not **T18b** or **T18c** were detected in this line. *THAA3* is presumably partially redundant as **T18a** was still detected in the *THAA3* mutant (*thaa3-ko*) (fig. S8, C and D). We

further showed that THAA3 could catalyze the formation of **T18b** when co-expressed with *THAS* and a characterized chain-length specific 14:0-acyl carrier protein (ACP) thioesterase gene from *Cuphea palustris*, *CpFatB2* in *N. benthamiana* (fig. S8A) (*32*), suggesting that there are as yet unidentified genes encoding chain-length specific ACP thioesterases involved in TFAE (**T18a-c**) biosynthesis. Besides thalianol, THAA3 could also act on other triterpenes, including arabidiol **A1** and its derivative **A2**, the PEN3 product **Tr1**, and marnerol **M2**, to introduce a palmityl or myristyl group depending on the type of fatty acyl CoA available (fig. S9). However, these products were not detected in *A. thaliana* Col-0 or *thaa3-oe* roots, possibly due to their very low abundance.

The promiscuity of enzymes encoded by the non-clustered [referred to as peripheral (*14*)] genes including *THAR1*, *THAR2* and *THAA2* and their strong co-expression with other divergent triterpene gene clusters in roots (Fig. 1B and Fig. 2) prompted us to perform combinatorial biosynthesis experiments in *N. benthamiana* using these genes and the arabidiol, marneral and tirucalladienol cluster TTS and CYP genes. Our results showed that THAR1, THAR2 and THAA2 could also act on arabidiol **A1**, 14-apo-arabidiol **A2** and tirucalladienol **Tr1** (fig. S10 and S11), but not marnerol **M2**, to give the corresponding C3 ketones, 3α-alcohols and acetates, respectively. Notably, arabidiol **A1** was fully converted to **A5** when *THAR1*, *THAR2* and *THAA2* were co-expressed together with *ABDS* and *CYP705A1* (Fig. 2 and fig. S12). **A5** is an *A. thaliana* root metabolite identified through comparative metabolomics analysis of the *A. thaliana* wild type (Col-0) with *ABDS* and *CYP705A1* mutants (*33*). Our results show that *THAR1*, *THAR2* and *THAA2* encode enzymes that can convert **A2** to **A3**, **A3** to **A4**, and **A4** to **A5**, respectively, thereby reconstituting the complete biosynthesis of **A5** in *N. benthamiana* (Fig. 2; fig. S12; table S17).

Metabolite analyses of root extracts of *A. thaliana* T-DNA insertion mutants of *THAR1* (*thar1-ko*), *THAR2* (*thar2-ko*) and *THAA2* (*thaa2-ko*) confirm that these three genes are indeed required for

the synthesis of thalianin **T10** in *A. thaliana*. **T10** was absent in root extracts from all three mutant genotypes and the pathway intermediates **T7** and **T17**, **T8**, and **T9**, respectively, accumulated instead (fig. S13). Additionally, *THAR2* and *THAA2* are also responsible for the biosynthesis of 14-apo-arabi-3α-yl acetate **A5** (referred hereafter as arabidin), since mutants *thar2-ko* and *thaa2-ko* lacked **A5** and instead accumulated **A3** and **A4**, respectively (fig. S14). **A5** was still detected in the root extracts from the *thar1-ko* line, indicating that *THAR1* is partially redundant in 14-apo-arabi-3α-yl acetate **A5** biosynthesis. Thus, these enzymes are involved in multiple pathways.

Triterpene biosynthesis affects *Arabidopsis* root microbiota assembly

Having discovered and reconstituted the complete biosynthetic pathways of thalianin **T10**, arabidin **A5**, and TFAEs **T18a-b** in *N. benthamiana*, we sought to investigate the potential biological role of this metabolic network. The thalianin **T10** and arabidin **A5** pathway genes are co-expressed in *A. thaliana* root epidermis and pericycle/stele, respectively (fig. S15) and thalianol pathway compounds **T1**, **T2**, **T9**, **T10** and arabidin **A5** could be detected in root surface wax extracts and **T9** and **T10** also in exudates of hydroponically grown seedlings (Fig. 3, D and E, fig. S16). Moreover, both the thalianin and arabidin pathway genes are upregulated upon methyl jasmonate treatment (fig. S17), suggesting that they may have a role in interactions with root microbes (*34*).

We selected mutants disrupted in the thalianin (**T10**), TFAE (**T18a-c**) and arabidin (**A5**) pathways (*thas-ko1*, *thas-ko2*, *thah-ko*, *thao-ko*, *thaa2-ko*, and *thaa2-crispr*) along with the wild type (Col-0) for root microbiota analyses (table S2). Comprehensive unbiased untargeted metabolomics analysis of whole root extracts, root surface extracts and root exudates from these mutant lines versus Col-0 suggests that the thalianin pathway compounds are the most affected of all root metabolites analysed, showing the largest fold changes across the respective metabolomes of the

mutants (figs. S18-22). We grew these lines in natural soil from Changping Farm in Beijing for 6 weeks under controlled experimental conditions. The roots were harvested and washed with PBS buffer to remove soil particles and loosely attached microbes prior to carrying out deep 16S rRNA gene sequencing of the root microbiota using previously established methodology (*3, 4, 35*). There were no obvious differences in root phenotypes between the wild type (Col-0) and mutant lines when the plants were harvested. However, we found that mutants affected in these pathways assembled different root microbiota when compared with the wild type (Col-0). Constrained principal ordination analysis (CPCoA) revealed differences in root microbiota between the wild type (Col-0) and mutant lines (Fig. 4A; 16.5% of total variance was explained by the plant genotypes, $P < 0.001$, PERMANOVA, tables S18-22). Pairs of independent mutant lines for *THAS* (*thas-ko1* and *thas-ko2*) and *THAA2* (*thaa2-ko* and *thaa2-cripsr*) each display similar metabolite defects (figs. S18-22) and have similar microbiota profiles and microbial diversity (Fig. 4A, fig. S23, and tables S23-26). Furthermore, all the pathway mutants tested showed similar root microbiota modulation patterns for Bacteroidetes (enrichment) and Deltaproteobacteria (depletion) compared to the *A. thaliana* Col-0 wild type at the phylum and OTU levels (Fig. 4B, figs. S24-25, tables S27-44), consistent with the fact that these genes operate in the same pathway for thalianin biosynthesis. Of the OTUs that were co-enriched/co-depleted in the mutants, 93% (28 out of 30) showed higher abundance in Col-0 roots than bulk soil, suggesting active selection of root bacteria by the *A. thaliana* plants (fig. S26, tables S45-46). These data indicate that the triterpene biosynthetic network that we have unveiled contributes to root microbiota assembly and establishment.

To understand whether and how this specialized metabolic network might modulate *A. thaliana*-specific root bacteria, we also compared the root bacterial profiles of the *A. thaliana* Col-0 and

mutant lines with those from the taxonomically distant species, rice (*36*) and wheat, also previously grown in the soil from Changping Farm on different occasions. Although the *A. thaliana*, rice and wheat samples differ from each other in many aspects including growth conditions, germination periods and climate conditions (see SI Materials and Methods), the starting inocula (soils) are very similar, sharing substantial overlap in total OTUs (67%, 2377 out of 3531, fig. S27, tables S47-49). We found that of the 494 OTUs specifically enriched in the root microbiota of *A. thaliana* Col-0 c.f. rice and wheat (represented in the overlap between the blue and orange circles, Fig. 4, C and D), 34% (170 out of 494, Fig. 4C, tables S50-55) were depleted and 18% (88 out of 494, Fig. 4D, tables S50-55) enriched in the root microbiota of the triterpene mutant lines compared to that of *A. thaliana* Col-0. Unlike *A. thaliana*, rice and wheat do not make thalianin, TFAEs or arabidin (fig. S28). Our results suggest that the specialized triterpene biosynthetic network that we have uncovered may contribute to enrichment of around a third of the *A. thaliana*-specific root bacteria present in the roots of the wild type line (Fig. 4C), while deterring another 18% of this bacterial population (Fig. 4D). The relatively larger number of OTUs depleted in the triterpene mutants (vs. Col-0) (a total of 380, pink circle, Fig. 4C) in comparison to the total number of OTUs that are enriched in the mutant lines vs. Col-0 (298, pink circle, Fig. 4D) further suggests that this triterpene biosynthetic network may play a more significant role in enriching *Arabidopsis*-specific root bacteria rather than repelling other bacteria.

Purified triterpenes selectively modulate root bacteria

To test whether triterpene pathway metabolites directly regulate root microbiota members, we isolated and identified bacteria from the roots of *A. thaliana* Col-0 plants grown in soil from the aforementioned Changping Farm by limiting dilution and barcoded sequencing (*35*). We selected a total of 19 bacterial strains belonging to 17 genera within three major bacterial phyla

(Proteobacteria, Actinobacteria and Firmicutes) that shared > 97% 16S rRNA gene similarity to the OTUs that showed differential abundances in the microbiota of *A. thaliana* wild type (Col-0), mutants, and soil (fig. S29, table S56). These bacterial strains were grown in liquid culture with a formulated cocktail of purified compounds that reflected the content and composition of the pathway metabolites in the roots of *A. thaliana* Col-0 (figs. S30-34). We found that most of the Proteobacteria strains tested proliferated faster in the presence of the *A. thaliana* triterpene mixture, whereas all five Actinobacteria strains were inhibited (Fig. 5A, figs. S30-34). The OTUs corresponding to these bacterial isolates (16 out of 19) showed consistent enrichment or depletion patterns in plant roots versus soil that corresponded with the growth promotion and inhibitory effects of the triterpene mixture, suggesting that the compounds tested contribute to the active selection of plant root bacteria (figs. S30-34). Moreover, the corresponding enrichment or depletion patterns for 10 bacterial genera (59% of the 17 tested) to which the metabolite-sensitive bacterial isolates belong were detected in the microbiota of at least one pathway mutant included in our microbiota analysis, i.e. the genera of bacteria for which growth is either promoted or inhibited by the triterpene cocktail mixture are depleted or enriched, respectively in the pathway mutants compared to the wild type (Col-0) (fig. S35, tables S57-63). Further tests of the sensitive strains with purified individual compounds revealed that pathway metabolites can selectively modulate the growth of bacteria, and that small structural differences between compounds can affect activities (fig. S36). For example, we found that Actinobacteria *Arthrobacter sp.* strain A224 was inhibited by compounds **T2**, **T9**, **T10**, and **A5** at 20 μM but not by the other triterpenes tested, whereas all compounds tested showed growth promoting effects on Gammaproteobacteira *Arenimonas sp.* strain A388 (Fig. 5, B and C). We also found that strain A475-1 (*Agromyces sp.*) has alcohol dehydrogenase activity and could selectively convert **T2** into **3-keto-T2** but not

T9/**T10** (Fig. 6, A and B; fig. S37A), whilst strain A215 (*Pseudomonas sp.*) has lipase activity and is able to cleave the TFAEs **T18a-c** to give **T1** and the corresponding fatty acids but not the acetate **T11** (Fig. 6, C and D and fig. S37, B to D). Moreover, strain A215 could utilize the cleavage product palmitic acid as a carbon source for proliferation (Fig. 6E). Such diverse and significant interaction patterns of root metabolites with taxonomically distinct root microbiota members suggest that the metabolites from this biosynthetic network alter the assembly of *A. thaliana* root microbiota. Taking the results of the root microbiota sequencing and *in vitro* bioassays together, we conclude that this triterpene biosynthetic network tunes the ecological niche for the assembly and maintenance of *A. thaliana* root microbiota.

Conclusion

The triterpene biosynthetic network described here has a latent capacity for the synthesis of over 50 root metabolites (Fig. 2). This is a relatively large number considering the total of ~300 non-volatile root metabolites that we detected in the polar methanolic (~220) and nonpolar ethyl acetate extracts (~85) of *A. thaliana* Col-0 roots under our experimental conditions (tables S64-65). This network originates from evolutionary divergent biosynthetic gene clusters coupled with cross-talk involving enzymes encoded by peripheral genes that service multiple biosynthetic pathways. These root metabolites can selectively regulate the growth of *Arabidopsis* root bacteria from different taxa by acting as antibiotics or proliferating agents. Biosynthesis of these specialized root triterpenes dynamically modulate a large portion [52%, 258 (170+88) out of 498 OTUs, Fig. 4, C and D] of the *A. thaliana*-specific bacteria, thereby shaping the *A. thaliana*-specific root microbiota. Triterpenes, a large and diverse groups of plant natural products (> 20,000 reported so far) (*12*), likely also sculpt the microbiota of other plant species, tailoring them to generate plant species-specific microbial communities, and may be useful for engineering root microbiota for sustainable

agriculture (*37-39*). It is tempting to speculate that metabolic diversification in the Plant Kingdom may provide a basis for communication and recognition that enables the shaping of microbial communities tailored to the needs of the host, and that this may in part explain the raison d' être for plant specialized metabolism.

Materials and methods

Detailed Materials and Methods can be found in the Supplementary Materials.

Plant materials

Arabidopsis thaliana accession Columbia (Col-0) was used as wild type in this study. All *A. thaliana* T-DNA insertion mutants were obtained from Nottingham Arabidopsis Stock Centre (NASC) unless otherwise stated. Homozygous mutants were identified by PCR-based genotyping using the primers listed in the table S5A. A 35S promoter-driven overexpression line of thalianol synthase (*thas-oe*) was generated previously (*15*). Overexpression lines for *THAA3* were generated by transforming *A. thaliana* Col-0 plants with *Agrobacterium tumefaciens* strain LBA4404 harboring the expression vector pMDC32 which contains the coding sequence of *THAA3* (*At3g51970*) under the control of the 35S promoter via floral dipping (*40*). Homozygous CRISPR knockout lines of *THAA1* (*At5g47980*) and *THAA2* (*At5g47950*) were generated by transforming *A. thaliana* Col-0 wild type with *A. tumefaciens* C58C1 harboring CRISPR/Cas9 constructs with specific single guide RNAs as previously described (*41*).

Cloning and transient expression

The coding sequences for *THAS* (*At5g48010*), *THAH* (*At5g48000*), *THAO* (*At5g47990*), *THAA1* (*At5g47980*) and *THAA2* (*At5g47950*), *THAR1* (*At3g29250*), *THAR2* (*At1g66800*), *THAA3*

(*At3g51970*), *ABDS* (*At4g15340*), *CYP705A1* (*At4g15330*), *At1g14960*, *At5g23840*, *At2g16005*, *At5g38030*, *At1g50560*, *At5g38020*, *MRO* (*At5g42580*), *CYP705A12* (*At5g42590*), *MRN* (*At5g42600*), *At5g12420*, *At5g55380*, *At3g49210*, *At1g54570*, *At3g49190*, *At3g26840* were amplified from a root cDNA library of the *A. thaliana* Col-0 accession (*26*). The coding sequences for *PEN3* (*At5g36150*), *CYP716A1* (*At5g36110*) and the truncated HMG CoA reductase (tHMGR) gene from oat had been cloned previously and were included in this study (*18, 27*). The coding sequence of *CpFatB2* from *Cuphea palustris* (accession no. U38189) was retrieved from NCBI and synthesized by Integrated DNA Technologies. These sequences were cloned into the pEAQ-HT expression vector for transient expression in *N. benthamiana* leaves as detailed in the supplementary materials.

Metabolite extraction and analysis

N. benthamiana leaves from transient expression experiments were harvested 5 days post infiltration, lyophilized, extracted with EtOAc and analyzed by GC-MS. Seedlings of *A. thaliana* Col-0 and mutants, rice and wheat were grown on Murashige–Skoog (¼ MS) agar plates or liquid media at 22^{o}C under short day conditions (8 h light/16 h dark) for 10 days. Roots were harvested from 10-day-old seedlings grown on agar plates, lyophilized, extracted with EtOAc and analyzed by GC-MS or extracted with MeOH and analyzed by LC-MS. Spent media and whole seedlings of 10-day-old *A. thaliana* grown in ¼ MS liquid media were harvested separately, lyophilized, extracted with MeOH and analyzed by LC-MS. Targeted and untargeted metabolomics analysis of GC-MS and LC-MS data were performed using Agilent MassHunter and XC-MS, respectively.

Production and isolation of thalianol and arabidiol derived metabolites

Compounds **T1-T10**, **3-keto-T2**, **T17**, ***cis*-T10**, ***cis*-T17**, **A1**, **A2** and **A4** were obtained by large-scale vacuum infiltration of *N. benthamiana* leaves with *A. tumefaciens* LBA4404 strains

harboring the corresponding expression constructs followed by extraction and purification. Compounds **3-keto-T1**, **T11**, **T18a-c**, **A3** and **A5** were chemically synthesized from the corresponding precursors.

Microbiota analysis

A. thaliana plants (Col-0 and triterpene mutants) were grown under controlled short-day conditions in the natural soil from Changping Farm (40°5'49''N, 116°24'44'' E, Beijing, China). Two wheat (Xiaoyan54 and Jing411) and rice (IR24 and Nipponbare) varieties were grown in Changping Farm under field conditions in 2016 and 2017, respectively. Roots from *A. thaliana*, wheat and rice were harvested six weeks post plantation in natural soil for 16S rRNA gene profiling. Data analysis was performed using QIIME 1.9.1 (*42*), USEARCH 10.0 (*43*) and in-house scripts. OTUs with differential abundance were identified with a negative binomial generalized linear model in the edgeR package with a foldchange threshold >1.2 (20). Venn diagrams were generated using the Venndiagram package (22).

***In vitro* bioassays**

Root-associated bacteria were isolated and identified from *A. thaliana* Col-0 plants grown in the aforementioned natural soil from Changping Farm with limiting dilution and barcoded sequencing (*35*). 19 bacterial isolates from diverse taxa that share greater than 97% 16S rRNA gene similarity with representative OTU sequences were selected for bioassay. The bacteria were cocultured with different formulated cocktails of triterpene mixtures or individual triterpenes in 1/10 TSB media and their growth (OD_{600}) monitored over 48 to 72h. Bacteria mediated transformation of triterpenes **T1**, **T11**, **T18a-c** was tested against all 19 bacteria whereas **T2**, **T9** and **T10** against sensitive bacterial strains A224, A475-1 and A479 only.

References and notes

1. L. Chae, T. Kim, R. Nilo-Poyanco, S. Y. Rhee, Genomic signatures of specialized metabolism in plants. *Science* **344**, 510-513 (2014).
2. R. L. Berendsen, C. M. J. Pieterse, P. A. H. M. Bakker, The rhizosphere microbiome and plant health. *Trends Plant Sci.* **17**, 478-486 (2012).
3. D. Bulgarelli, M. Rott, K. Schlaeppi, E. Ver Loren van Themaat, N. Ahmadinejad, F. Assenza, P. Rauf, B. Huettel, R. Reinhardt, E. Schmelzer, J. Peplies, F. O. Gloeckner, R. Amann, T. Eickhorst, P. Schulze-Lefert, Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* **488**, 91 (2012).
4. D. S. Lundberg, S. L. Lebeis, S. H. Paredes, S. Yourstone, J. Gehring, S. Malfatti, J. Tremblay, A. Engelbrektson, V. Kunin, T. G. d. Rio, R. C. Edgar, T. Eickhorst, R. E. Ley, P. Hugenholtz, S. G. Tringe, J. L. Dangl, Defining the core Arabidopsis thaliana root microbiome. *Nature* **488**, 86 (2012).
5. C. Nguyen, Rhizodeposition of organic C by plants: mechanisms and controls. *Agronomie* **23**, 375-396 (2003).
6. K. Zhalnina, K. B. Louie, Z. Hao, N. Mansoori, U. N. da Rocha, S. Shi, H. Cho, U. Karaoz, D. Loqué, B. P. Bowen, M. K. Firestone, T. R. Northen, E. L. Brodie, Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nature Microbiology* **3**, 470-480 (2018).
7. I. A. Stringlis, K. Yu, K. Feussner, R. de Jonge, S. Van Bentum, M. C. Van Verk, R. L. Berendsen, P. A. H. M. Bakker, I. Feussner, C. M. J. Pieterse, MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. *Proc. Natl. Acad. Sci.* **115**, E5213-E5222 (2018).

8. S. L. Lebeis, S. H. Paredes, D. S. Lundberg, N. Breakfield, J. Gehring, M. McDonald, S. Malfatti, T. Glavina del Rio, C. D. Jones, S. G. Tringe, J. L. Dangl, Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349**, 860-864 (2015).

9. K. Papadopoulou, R. E. Melton, M. Leggett, M. J. Daniels, A. E. Osbourn, Compromised disease resistance in saponin-deficient plants. *Proc. Natl. Acad. Sci.* **96**, 12923-12928 (1999).

10. A. G. Pacheco, A. F. C. Alcântara, G. M. Corrêa, V. G. C. Abreu, *Relationships Between Chemical Structure and Activity of Triterpenes Against Gram-Positive and Gram-Negative Bacteria*. (INTECH Open Access Publisher, 2012).

11. J. M. Augustin, V. Kuzina, S. B. Andersen, S. Bak, Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry* **72**, 435-457 (2011).

12. R. Thimmappa, K. Geisler, T. Louveau, P. O'Maille, A. Osbourn, Triterpene Biosynthesis in Plants. *Annu. Rev. Plant Biol.* **65**, 225-257 (2014).

13. A. M. Boutanaev, A. Osbourn, Multigenome analysis implicates miniature inverted-repeat transposable elements (MITEs) in metabolic diversification in eudicots. *Proc. Natl. Acad. Sci.* **115**, E6650-E6658 (2018).

14. H.-W. Nützmann, A. Huang, A. Osbourn, Plant metabolic clusters – from genetics to genomics. *New Phytol.* **211**, 771-789 (2016).

15. B. Field, A. Osbourn, Metabolic diversification—independent assembly of operon-like gene clusters in different plants. *Science* **320**, 543-547 (2008).

16. B. Field, A.-S. Fiston-Lavier, A. Kemen, K. Geisler, H. Quesneville, A. Osbourn, Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc. Natl. Acad. Sci.* **108**, 16116-16121 (2011).

17. R. Sohrabi, J.-H. Huh, S. Badieyan, L. H. Rakotondraibe, D. J. Kliebenstein, P. Sobrado, D. Tholl, In planta variation of volatile biosynthesis: An alternative biosynthetic route to the formation of the pathogen-induced volatile homoterpene DMNT via triterpene degradation in Arabidopsis roots. *The Plant Cell* **27**, 874-890 (2015).

18. A. M. Boutanaev, T. Moses, J. Zi, D. R. Nelson, S. T. Mugford, R. J. Peters, A. Osbourn, Investigation of terpene diversification across multiple sequenced plant genomes. *Proc. Natl. Acad. Sci.* **112**, E81-E88 (2015).

19. Q. Xiong, W. K. Wilson, S. P. T. Matsuda, An Arabidopsis oxidosqualene cyclase catalyzes iridal skeleton formation by Grob fragmentation. *Angew. Chem. Int. Ed.* **45**, 1285-1288 (2006).

20. G. C. Fazio, R. Xu, S. P. T. Matsuda, Genome mining to identify new plant triterpenoids. *J. Am. Chem. Soc.* **126**, 5678-5679 (2004).

21. T. Xiang, M. Shibuya, Y. Katsube, T. Tsutsumi, M. Otsuka, H. Zhang, K. Masuda, Y. Ebizuka, A new triterpene synthase from Arabidopsis thaliana produces a tricyclic triterpene with two hydroxyl groups. *Org. Lett.* **8**, 2835-2838 (2006).

22. P. Morlacchi, W. K. Wilson, Q. Xiong, A. Bhaduri, D. Sttivend, M. D. Kolesnikova, S. P. T. Matsuda, Product profile of PEN3: The last unexamined oxidosqualene cyclase in Arabidopsis thaliana. *Org. Lett.* **11**, 2627-2630 (2009).

23. S. Lodeiro, Q. Xiong, W. K. Wilson, M. D. Kolesnikova, C. S. Onak, S. P. T. Matsuda, An oxidosqualene cyclase makes numerous products by diverse mechanisms: A

challenge to prevailing concepts of triterpene biosynthesis. *J. Am. Chem. Soc.* **129**, 11213-11222 (2007).

24. D. A. Castillo, M. D. Kolesnikova, S. P. T. Matsuda, An effective strategy for exploring unknown metabolic pathways by genome mining. *J. Am. Chem. Soc.* **135**, 5885-5894 (2013).

25. J. C. D'Auria, Acyltransferases in plants: a good time to be BAHD. *Curr. Opin. Plant Biol.* **9**, 331-340 (2006).

26. A. C. Huang, S. A. Kautsar, Y. J. Hong, M. H. Medema, A. D. Bond, D. J. Tantillo, A. Osbourn, Unearthing a sesterterpene biosynthetic repertoire in the Brassicaceae through genome mining reveals convergent evolution. *Proc. Natl. Acad. Sci.* **114**, E6005-E6014 (2017).

27. J. Reed, M. J. Stephenson, K. Miettinen, B. Brouwer, A. Leveau, P. Brett, R. J. M. Goss, A. Goossens, M. A. O'Connell, A. Osbourn, A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng.* **42**, 185-193 (2017).

28. A. C. Huang, Y. J. Hong, A. D. Bond, D. J. Tantillo, A. Osbourn, Diverged plant terpene synthases reroute the carbocation cyclization path towards the formation of unprecedented 6/11/5 and 6/6/7/5 sesterterpene scaffolds. *Angew. Chem. Int. Ed.* **57**, 1291-1295 (2018).

29. T. Obayashi, Y. Aoki, S. Tadaka, Y. Kagaya, K. Kinoshita, ATTED-II in 2018: A plant coexpression database based on investigation of the statistical property of the mutual rank index. *Plant Cell Physiol.* **59**, e3-e3 (2018).

30. D. Ma, Z. Wang, C. N. Merrikh, K. S. Lang, P. Lu, X. Li, H. Merrikh, Z. Rao, W. Xu, Crystal structure of a membrane-bound O-acyltransferase. *Nature* **562**, 286-290 (2018).

31. Q. Chen, L. Steinhauer, J. Hammerlindl, W. Keller, J. Zou, Biosynthesis of phytosterol esters: Identification of a sterol O-scyltransferase in Arabidopsis. *Plant Physiol.* **145**, 974-984 (2007).

32. K. Dehesh, P. Edwards, T. Hayes, A. M. Cranmer, J. Fillatti, Two novel thioesterases sre key determinants of the bimodal distribution ofacyl chain length of Cuphea palustris seed oil. *Plant Physiol.* **110**, 203-210 (1996).

33. R. Sohrabi, T. Ali, L. Harinantenaina Rakotondraibe, D. Tholl, Formation and exudation of non-volatile products of the arabidiol triterpenoid degradation pathway in Arabidopsis roots. *Plant Signaling & Behavior* **12**, e1265722 (2017).

34. L. Hu, C. A. M. Robert, S. Cadot, X. Zhang, M. Ye, B. Li, D. Manzo, N. Chervet, T. Steinger, M. G. A. van der Heijden, K. Schlaeppi, M. Erb, Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nature Communications* **9**, 2738 (2018).

35. Y. Bai, D. B. Müller, G. Srinivas, R. Garrido-Oter, E. Potthoff, M. Rott, N. Dombrowski, P. C. Münch, S. Spaepen, M. Remus-Emsermann, B. Hüttel, A. C. McHardy, J. A. Vorholt, P. Schulze-Lefert, Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* **528**, 364 (2015).

36. J. Zhang, N. Zhang, Y.-X. Liu, X. Zhang, B. Hu, Y. Qin, H. Xu, H. Wang, X. Guo, J. Qian, W. Wang, P. Zhang, T. Jin, C. Chu, Y. Bai, Root microbiota shift in rice correlates with resident time in the field and developmental stage. *Science China Life Sciences* **61**, 613-621 (2018).

37. B. O. Oyserman, M. H. Medema, J. M. Raaijmakers, Road MAPs to engineer host microbiomes. *Curr. Opin. Microbiol.* **43**, 46-54 (2018).

38. Y. Dessaux, C. Grandclément, D. Faure, Engineering the rhizosphere. *Trends Plant Sci.* **21**, 266-278 (2016).

39. H. Toju, K. G. Peay, M. Yamamichi, K. Narisawa, K. Hiruma, K. Naito, S. Fukuda, M. Ushio, S. Nakaoka, Y. Onoda, K. Yoshida, K. Schlaeppi, Y. Bai, R. Sugiura, Y. Ichihashi, K. Minamisawa, E. T. Kiers, Core microbiomes for sustainable agroecosystems. *Nature Plants* **4**, 247-257 (2018).

40. C. S. J., B. A. F., Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* **16**, 735-743 (1998).

41. A. Ritter, S. Iñigo, P. Fernández-Calvo, K. S. Heyndrickx, S. Dhondt, H. Shi, L. De Milde, R. Vanden Bossche, R. De Clercq, D. Eeckhout, M. Ron, D. E. Somers, D. Inzé, K. Gevaert, G. De Jaeger, K. Vandepoele, L. Pauwels, A. Goossens, The transcriptional repressor complex FRS7-FRS12 regulates flowering time and growth in Arabidopsis. *Nature Communications* **8**, 15235 (2017).

42. J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335-336 (2010).

43. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461 (2010).

44. BIG Data Center Members, Database Resources of the BIG Data Center in 2018. *Nucleic Acids Res.* **46**, D14-D20 (2017).

45. D. Winter, B. Vinegar, H. Nahal, R. Ammar, G. V. Wilson, N. J. Provart, An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS One* **2**, e718 (2007).

Acknowledgments. J. Pollier and P. Fernandez-Calvo are acknowledged for their support of Y-C.B. C. Owen is acknowledged for initial testing of the thalianol cluster genes. **Funding:** This work has been supported by the National Institutes of Health Genome to Natural Products Network award U101GM110699 (A.O., A.C.H.), the 'Strategic Priority Research Program' of the Chinese Academy of Sciences (XDB11020700) (Y.B.), the international cooperation and exchanges NSFC Grant No. 31761143017 (Y.B.), the Centre of Excellence for Plant and Microbial Sciences (CEPAMS), established between the John Innes Centre and the Chinese Academy of Sciences and funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC) and the Chinese Academy of Sciences (A.O., Y.B.), the Priority Research Program of the Chinese Academy of Sciences (QYZDB-SSW-SMC021) (Y.B.), the European Community's Seventh Framework Program (FP7/2007–2013) under grant agreement 613692 (TriForC) (A.O., A.G.), the joint Engineering and Physical Sciences Research Council/ BBSRC-funded OpenPlant Synthetic Biology Research Centre grant BB/L014130/1 (H-W.N., A.O.) and the Research Foundation Flanders with a research project grant to A.G (G008417N). . A.C.H. is supported by a European Commission Marie Skłodowska-Curie Individual Fellowship (H2020-MSCA-IF-EF-ST-702478-TRIGEM). H-W.N. is currently supported by a Royal Society University Research Fellowship (UF160138). A.O.'s laboratory is funded by the UK Biotechnological and Biological Sciences Research Council (BBSRC) Institute Strategic

Programme Grant ‘Molecules from Nature’ (BB/P012523/1) and the John Innes Foundation. Y-C.B. is supported by a China Scholarship Council (CSC) Ph.D. scholarship.

Author contributions. A.C.H., Y. B. and A.O. conceived and designed the project. A.C.H. discovered and characterized the biosynthetic network, performed bacterial growth assay and coordinated the project; T.J. grew plants in natural soils, harvested roots, prepared the 16S amplicon library for sequencing, isolated *A. thaliana* root bacteria and performed bacterial growth assays; Y.X.L. performed bioinformatics analysis on microbiota sequencing results; Y-C.B. generated the homozygous *thaa1-crispr* and *thaa2-crispr* lines; J.R. cloned the thalianol and marneral cluster genes; B.Q. grew and harvested the wheat samples for microbiota analysis. A.C.H., T.J., Y.X.L., H.N., A.G. and Y.B. analyzed data; A.C.H., T.J., Y.B. and A.O. wrote the manuscript with contributions from other authors.

Competing interest: The author declare that they have no competing interest.

Data and materials availability:

Raw microbiota sequencing data reported in this paper have been deposited in the Genome Sequence Archive in Beijing Institute of Genomics (BIG) Data Center (*44*), Chinese Academy of Sciences under accession number, PRJCA001296 that are public accessible at http://bigd.big.ac.cn/gsa. Scripts employed in the microbiota analyses are available under the following link: https://github.com/microbiota/Huang2019SCIENCE.

Supplementary Materials:

Materials and Methods

Figures S1-S64

Tables S1-S65

References (*46-60*)

Figure legends:

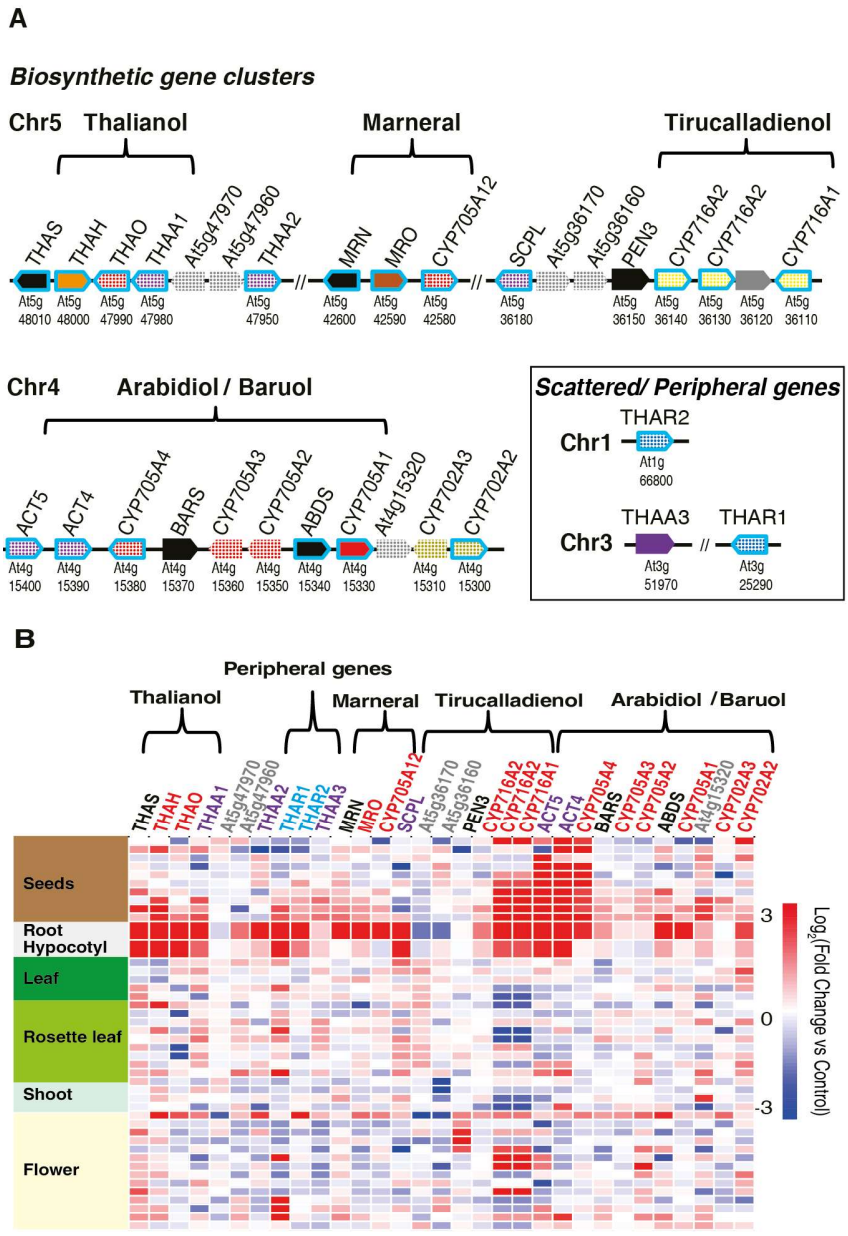

Fig. 1. Identification of a root expressed triterpene biosynthetic network in *A. thaliana*. (**A**) The four *A. thaliana* triterpene biosynthetic gene clusters and peripheral genes. Genes encoding

previously characterized enzymes are shown in solid color and uncharacterized ones are dotted. Black, triterpene synthases; red, CYP705 family; orange, CYP708 family; brown, CYP71 family; yellow, CYP716 family; dark yellow, CYP702 family; purple, acyltransferases (ACTs); blue, alcohol dehydrogenase (ALDHs); grey, non-biosynthetic genes. The scattered/peripheral genes identified as part of this network are also shown. Co-expressed genes have light blue borders. PEN3 was previously shown to be co-expressed with other cluster genes in roots (*18*). (**B**) Expression profiles of the triterpene cluster genes and additional scattered (peripheral) biosynthetic genes identified as part of this study in different *A. thaliana* tissues. The heatmap was generated using microarray expression data from the eFP browser (*45*). For clarity, expression profiles for the same tissue across different developmental stages are labelled with the tissue names only. A more detailed heatmap can be found in fig. S1.

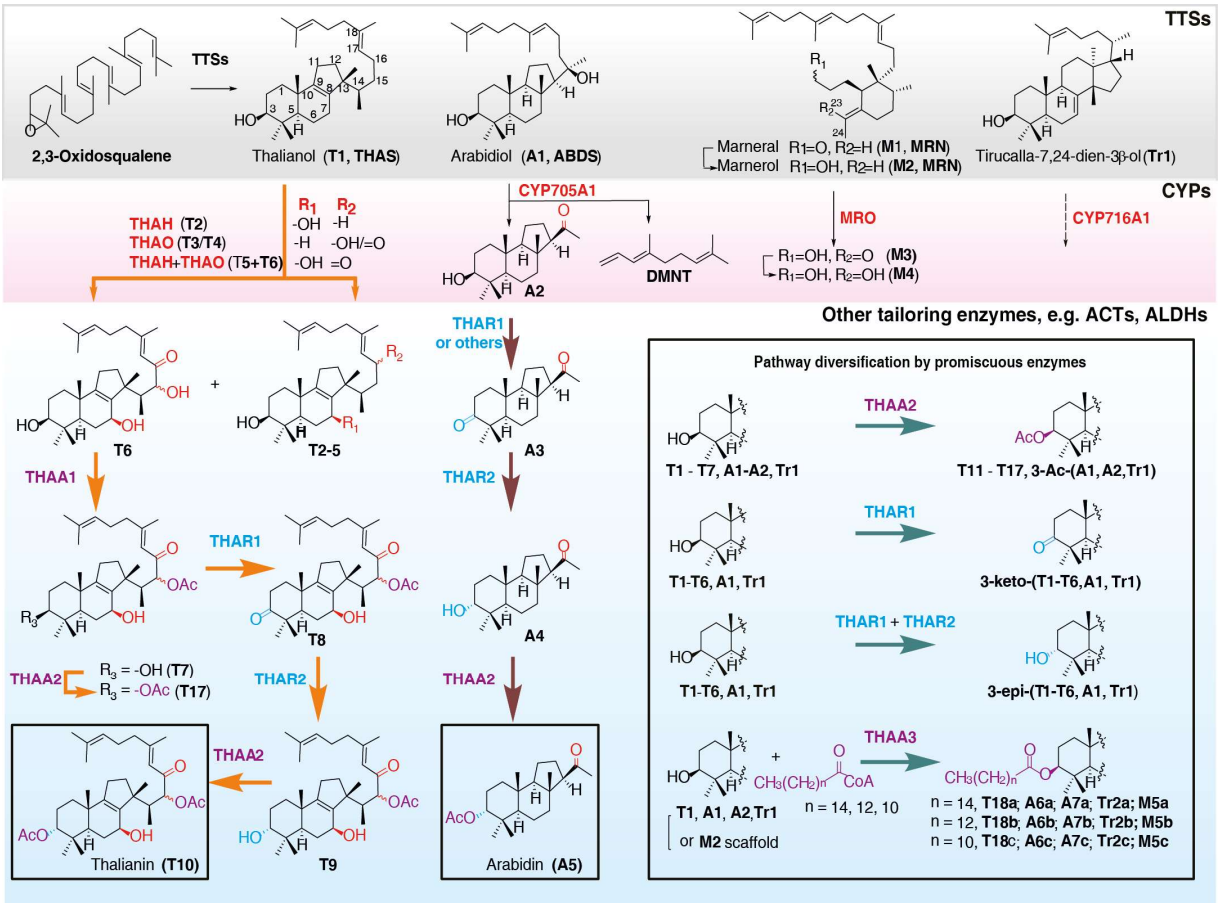

Fig. 2. **The triterpene biosynthetic network in *A. thaliana* roots**. Transformations catalyzed by TTSs are gray shaded, clustered CYPs red shaded, and other tailoring enzymes blue shaded. Chemical diversification of pathway scaffolds and intermediates by promiscuous acyltransferases and alcohol dehydrogenases is highlighted in the black box, bottom right. DMNT, (*E*)-4,8-dimethyl-1,3,7-nonatriene. Black arrows, known transformations; orange and burgundy red arrows, newly characterized steps in the thalianin and arabidin pathways, respectively; Cyan arrows, pathway diversification by promiscuous enzymes; dotted black arrow, unknown transformation. The enzymes for each step are color-coded to match the classes of enzyme shown in Fig. 1A, with the exception that all CYPs are in red.

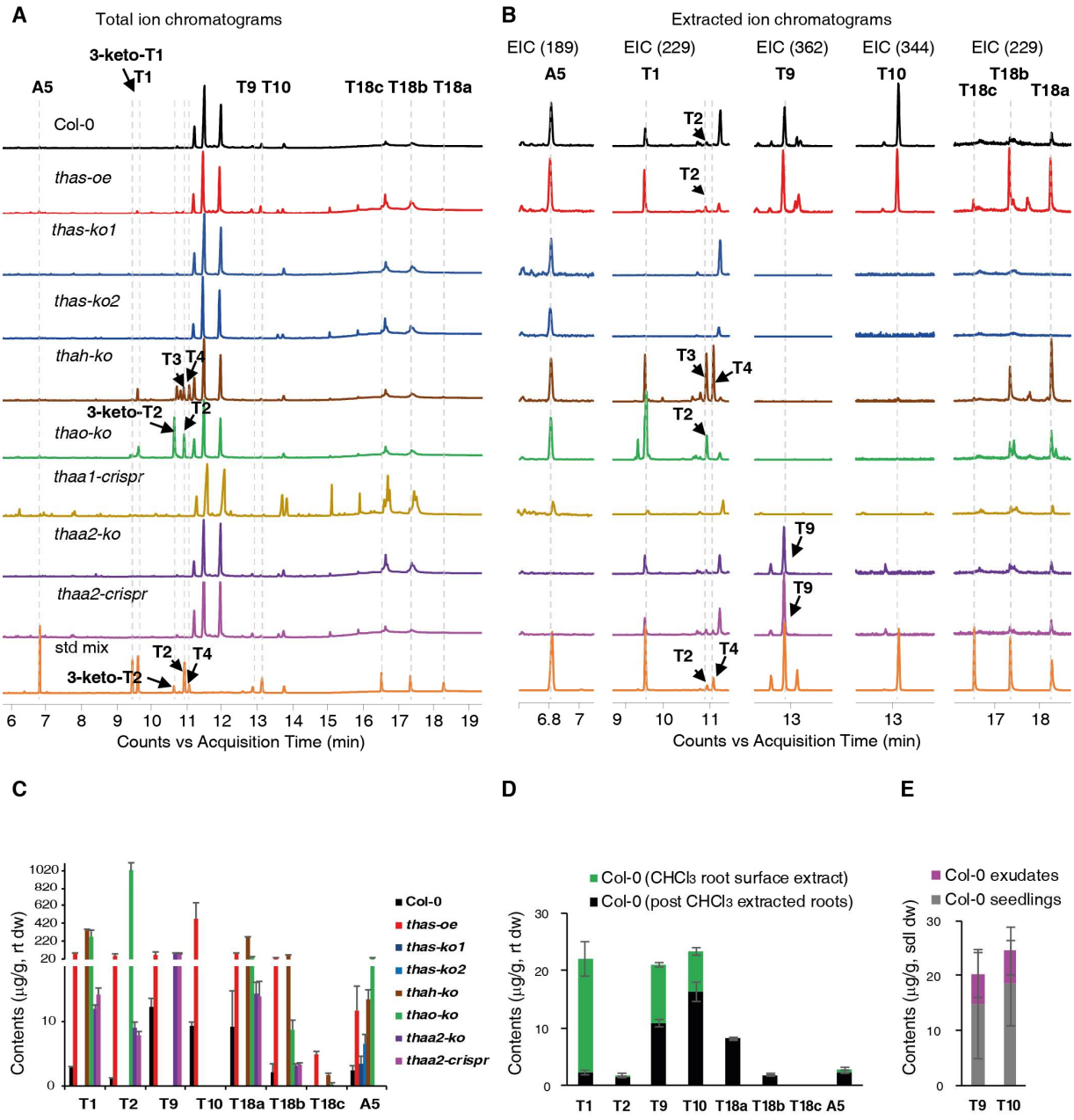

Fig. 3. Metabolite analysis of roots of different *Arabidopsis* lines. (**A**) Comparative GC-MS total ion chromatograms (TICs) of root extracts of the *A. thaliana* wild type Col-0 (black), *thas-oe* (red), *thas-ko1* (dark blue), *thas-ko2* (blue), *thah-ko* (brown), *thao-ko* (green), *thaa1-crispr* (dark yellow), *thaa2-ko* (purple), *thaa2-crispr* (pink) and an authentic standard mix containing **T1**, **3-keto-T1**, **T2**, **3-keto-T2**, **T4**, **T9**, **T10**, **T18a-c** and **A5** (orange). (**B**) Comparative extracted ion

chromatograms (EICs) from Fig. 2A for characteristic fragments of the mass spectra of **A5** (189), **T1-4** (229), **T9** (362), **T10** (344) and **T18a-c** (229). (**C**) Triterpene pathway metabolites detected in whole root extracts from wild type and mutant lines. Col-0, wild type; *thas-oe*, *thas-ko1*, *thas-ko2*, *thah-ko*, *thao-ko*, *thaa2-ko* and *thaa2-crispr*, mutants for *THAS*, *THAH*, *THAO* and *THAA2* (see SI Materials and Methods). (**D**) Triterpenoids detected in chloroform ($CHCl_3$) extracts from the surfaces of fresh roots of *A. thaliana* wild type Col-0 seedlings and subsequent ethyl acetate extracts of dry and powdered chloroform-extracted roots. All lines used in qualitative and quantitative GC-MS analysis (Fig. 2A-2D) were grown on ¼ MS agar plates at 22°C (8 h dark/16 h light cycle) for 10 days. (**E**) Determination of compounds **T9** and **T10** by LC-MS in root exudates of Col-0 seedlings grown hydroponically in liquid ¼ MS after 10 days. Error bars represent standard deviation of three biological replicates.

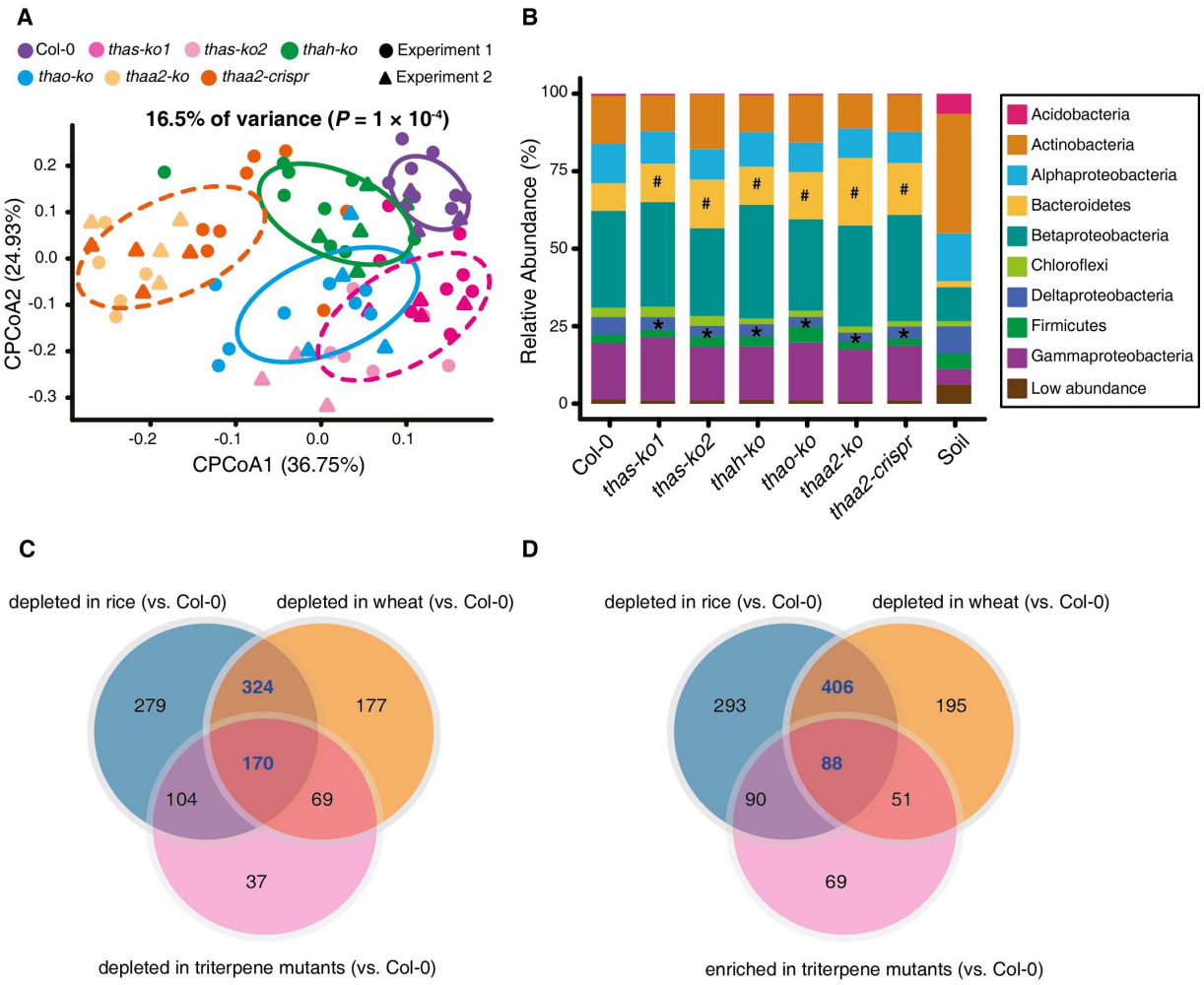

Fig. 4. Modulation of specific root bacterial taxa in triterpene pathway mutants. (**A**) Constrained principal ordination analysis (CPCoA) of Bray-Curtis dissimilarity showing plant genotype effects. Total number of individual plants used for analyses: Col-0 (n = 12), *thas-ko1* (n = 12), *thas-ko2* (n = 9), *thah-ko* (n = 14), *thao-ko* (n = 13), *thaa2-ko* (n = 9), *thaa2-crispr* (n = 12). Biological replicates (individual plants) from two independent experiments (experiment 1 and 2) are indicated by dots and triangles, respectively. Ellipses include 68% of samples from each genotype. (**B**) Phylum distribution of the root microbiota compositions of the tested *A. thaliana* genotypes**.** As the relative abundance of Proteobacteria is more than 50%, bacteria in this phylum

are shown at the class level. # indicates Bacteroidetes significantly higher than that in Col-0 roots at $P < 0.05$; * indicates Alphaproteobacteria and Deltaproteobacteria significantly lower than that in Col-0 at $P < 0.05$. (**C**, **D**) Venn diagrams showing significant overlap of OTUs depleted (**C**) or enriched (**D**) in the root microbiota of *A. thaliana* triterpene mutant lines c.f. the wild type (Col-0) (pink circles) with those depleted in the root microbiota of rice (blue circles) and wheat (orange circles) versus the *A. thaliana* Col-0 wild type. The OTU numbers specifically enriched in the root microbiota of *A. thaliana* Col-0 c.f. rice and wheat are highlighted in blue and bold in the Venn diagram overlaps.

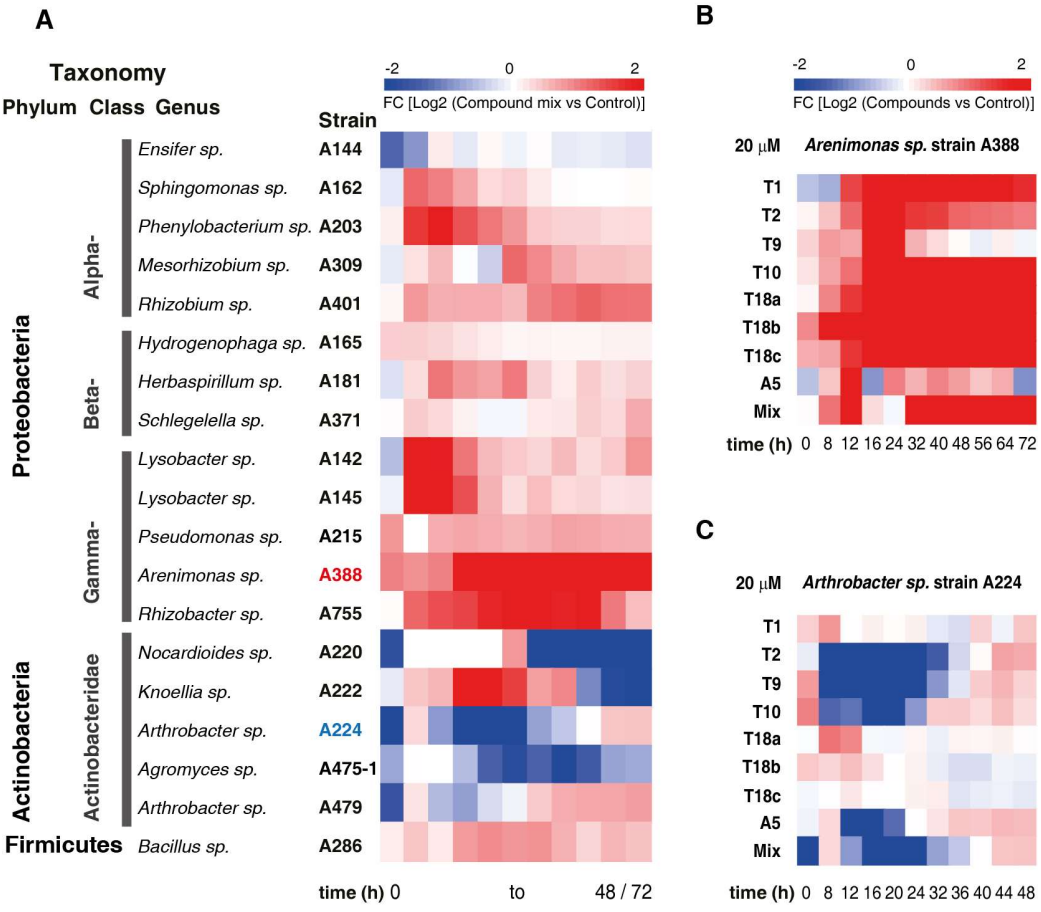

Fig. 5. Effects of pathway metabolites on the growth of isolated *A. thaliana* root-associated bacteria and bacterium-mediated chemical transformations. (**A**) Growth modulation activity

of compound mixture **Mix** (10 µM **T1**, 5 µM **T2**, 20 µM **T9**, 20 µM **T10**, 10 µM **T18a**, 5 µM **T18b**, 1 µM **T18c** and 10 µM **A5**) against the 19 strains of *A. thaliana* root bacteria from different taxa. The heatmap shows log_2 fold change of individual strain treated with **Mix** vs. control (ethanol only) over 48 or 72 hours. The corresponding graphical growth curves are depicted in figs. S30-34. (**B**) Heatmap showing the log_2fold change in cell density (OD_{600}) of *Arenimonas sp.* strain A388 treated with 8 different purified compounds (**T1**, **T2**, **T9**, **T10**, **T18a**, **T18b**, **T18c**, and **A5**, respectively) at 20 µM and **Mix** over 72 h. All compounds were dissolved in ethanol. Control, 0 mM, ethanol. (**C**) Heatmap showing the log_2fold change in cell density (OD_{600}) of *Arthrobacter sp.* strain A224 treated with 8 different purified compounds (**T1**, **T2**, **T9**, **T10**, **T18a**, **T18b**, **T18c**, and **A5**, respectively) at 20 µM and **Mix** over 48 h. All compounds were dissolved in ethanol. Control, 0 mM, ethanol.

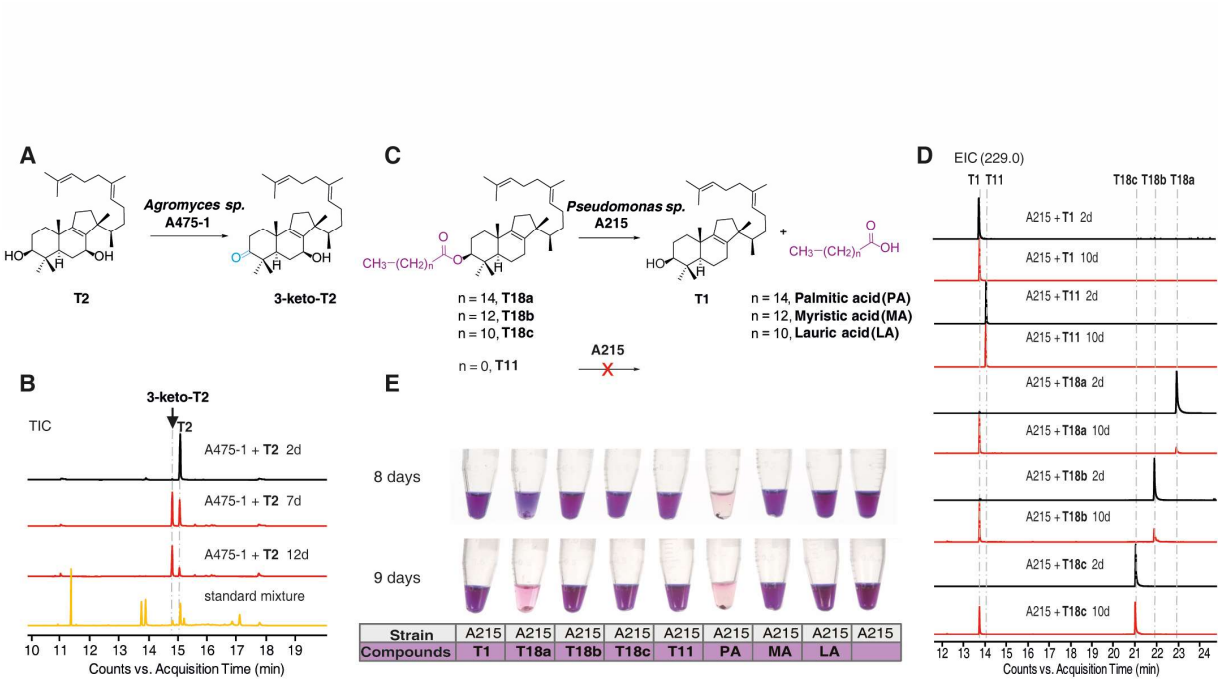

Fig. 6. Bacterium-mediated transformations of triterpenes. (**A**) Conversion of thaliandiol **T2** to **3-keto-T2** by *Agromyces sp.* strain A475-1 as shown by (**B**) comparative GC-MS TICs of EtOAc extracts of bacterial cultures of supplemented with **T2** on the 2nd, 7th and 10th day, respectively. (**C**) Selective conversion of TFAEs **T18a-c**, but not **T11** to **T1** by *Pseudomonas sp.* strain A215 as shown by (**D**) comparative GC-MS EICs (*m/z* = 229) of EtOAc extracts of bacterial strain A215 cultures supplemented with **T1**, **T11** and **T18a-c** on the 2nd and 10th day, respectively. (**E**) *Pseudomonas sp.* strain A215 could utilize palmitic acid as carbon source as shown by the proliferation of A215 in minimal salt media supplemented with 0.5 mM (in ethanol) of **T18a** or palmitic acid (**PA**) but not in those with **T1**, **T18b**, **T18c**, **T11**, myristic acid (**MA**), lauric acid (**LA**) and control. The minimal salt media contain 0.3 mg/ml resazurin as indicator of bacterial growth (from purple blue to pink).